Theor Appl Genet (2002) 104:414–421

X.-Y. Lou · J. Zhu

Analysis of genetic effects of major genes and polygenes on quantitative traits

I. Genetic model for diploid plants and animals

Received: 11 November 2000 / Accepted: 2 May 2001

Abstract A genetic model was proposed to simultaneously investigate genetic effects of both polygenes and several single genes for quantitative traits of diploid plants and animals. Mixed linear model approaches were employed for statistical analysis. Based on two mating designs, a full diallel cross and a modified diallel cross including F_2, Monte Carlo simulations were conducted to evaluate the unbiasedness and efficiency of the estimation of generalized least squares (GLS) and ordinary least squares (OLS) for fixed effects and of minimum norm quadratic unbiased estimation (MINQUE) and Henderson III for variance components. Estimates of MINQUE (1) were unbiased and efficient in both reduced and full genetic models. Henderson III could have a large bias when used to analyze the full genetic model. Simulation results also showed that GLS and OLS were good methods to estimate fixed effects in the genetic models. Data on *Drosophila melanogaster* from Gilbert were used as a worked example to demonstrate the parameter estimation.

Keywords Single gene · Polygenes · Genetic model for diploid organism · *GE* interaction · *Esterase* 6 gene in *Drosophila melanogaster*

Introduction

Many quantitative traits are affected by one or more major genes as well as by polygenes in some genetic material (Falconer 1981; Mather and Jinks 1982). In these cases separation of the effects of major genes and polygenes is of great importance for understanding their expression in genetics and for evaluating their utilization in breeding. Since the inheritance of these traits is different from that of quantitative traits in a traditional sense, the classic analysis methods are no longer valid for them. There need to be appropriate methods for obtaining correct genetic conclusions.

As reviewed by Gilbert (1985b), single locus effects (e.g. additive and dominance) on quantitative traits are often estimated by comparisons between genetic entries with different genotypes at this locus if general genetic differences due to residual background genotypes can be controlled or accounted for. From those comparisons, however, the magnitude of genetic effects resulting from polygenes cannot be rigorously determined, and their relative importance to trait difference may remain unknown. Mo (1993a, b) and Mo and Xu (1994) proposed methods for identifying major gene genotypes and for estimating the effects of a major gene and polygenes from generation-populations derived from a cross between two pure parental lines. Mixture distribution models and the likelihood-based analysis technique have also widely been used to estimate the genetic parameters based on populations descended from a cross between a pair of inbred lines (Fernando et al. 1994; Jiang et al. 1994; Loisel et al. 1994; Jiang and Liu 1995; Jiang and Mo 1995; Jiang et al. 1995; Wang and Gai 1997; Gai and Wang 1998). But the generality of the major gene and polygene estimates depend on the absence of interaction between the major genes and the polygenes, as well as on the representiveness of the initial cross for relevant polygenic constitutions. Furthermore, the estimates are only relative to a specific genetic background from a fixed set of materials, and generalized conclusions cannot be drawn for the population of interest. Elkind and Cahaner (1986) developed a mixed model for effects of single gene, polygenes and their interaction on quantitative traits. Nevertheless, estimation is limited to the total genetic effects of a single gene and polygenes. The nested (Design I) and the factorial (Design II) designs (Comstock and Robinson 1952), and diallel mating designs (Yates 1947, Hayman 1954a, b; Griffing 1956; Gilbert 1958; Eberhart and Gardner 1966; Gardner and Eberhart

Communicated by H.H. Geiger

X.-Y. Lou · J. Zhu (✉)
Department of Agronomy, Zhejiang University,
Hangzhou, Zhejian 310029, PRC
e-mail: jzhu@zju.edu.cn
Tel.: +86-571-6971444, Fax: +86-571-6971117

(1966), are most used when genetic properties of a reference population or multiple lines are studied for polygenic variation. Gilbert (1985a, b) suggested a diallel method for estimating single gene effects and polygenic background effects. This method is incapable of estimating the interaction between the single gene and polygenes. Moreover, it cannot deal with unbalanced data and the genetic design required is very strict.

In general, the nested and factorial designs can be viewed as a subset of diallel designs with some missing entries. In the present research, a genetic model is proposed for quantitative traits influenced by both major genes and polygenes from modified diallel designs in diploid plants or animals. In consideration of the complication of the genetic model, mixed linear model approaches are suggested for statitistical analysis. Monte Carlo simulations are conducted to evaluate estimation methods. Data for mating speed, copulation duration developmental time, and esterase 6 enzyme activity in *Drosophila melanogaster* from Gilbert (1985a) are used as a worked example.

Model and methodology

If the genetic difference of a quantitative trait is jointly controlled by polygenes and several single genes, which refer to any identifiable locus not confined to a major gene with larger effect, the phenotypic mean (y_{hijkl}) of mating-type k of the combination from maternal parent i and paternal parent j in block l within environment h can be expressed by a linear model as follows:

$$\begin{aligned} y_{hijkl} &= \mu + E_h + G_{ijk} + GE_{hijk} + B_{l(h)} + e_{hijkl} \\ &= \mu + E_h + Gs_{ijk} + Gp_{ijk} + Gsp_{ijk} + GsE_{hijk} + GpE_{hijk} \\ &\quad + GspE_{hijk} + B_{l(h)} + e_{hijkl}, \end{aligned} \tag{1}$$

where μ is the fixed population mean; E_h is the effect of environment h (e.g. year, location, etc.), fixed or random (determined by context of data), and is usually random in genetic experiments since genetic expression in all environments is of interest; $B_{l(h)} \sim (0, \sigma_B^2)$ is the effect of block l within environment h; $e_{hijkl} \sim (0, \sigma_e^2)$ is the residual effect; G_{ijk} is the genotypic value or total genetic effect, which can be further partitioned into three components with Gs_{ijk} attributable to single nuclear genes, Gp_{ijk} to background polygenes, and Gsp_{ijk} to the interaction effect between single genes and polygenes; GE_{hijk} is the total interaction effect between genotype G_{ijk} and environment E_h, which also consist of three components with GsE_{hijk} the interaction effect between Gs_{ijk} and E_h, GpE_{hijk} the interaction effect between Gp_{ijk} and E_h, and $GspE_{hijk}$ the interaction effect between Gsp_{ijk} and E_h. If there is no interaction between single genes and polygenes, the phenotypic value can be expressed as a reduced model without the terms Gsp_{ijk} and $GspE_{hijk}$.

Assuming (1) regular segregation, (2) autosomal inheritance for variation attributable to nuclear genes, (3) absence of maternal effects, (4) no epistatic effects between single genes and between polygenes, and (5) constant transmission of cytoplasmic genes through the maternal parent, then Cockerham's (1980) general genetic model can be extended by adding the cytoplasmic effect but excluding epistatic effects to represent the genotypic value associated with polygenes ($x, y \in \{i, j\}$), i.e.:

$$\begin{aligned} Gp_{ijk} &= \alpha(k)_i A_i + \alpha(k)_j A_j + \delta(k)_{ii} D_{ii} + \delta(k)_{jj} D_{jj} + \delta(k)_{ij} D_{ij} + \gamma(k)_i C_i + \gamma(k)_j C_j \\ &= \sum_x \alpha(k)_x A_x + \sum_{x \le y} \delta(k)_{xy} D_{xy} + \sum_x \gamma(k)_x C_x, \end{aligned} \tag{2}$$

where i represents the polygenic ancestry source for nuclear genes and for cytoplasmic genes from which parent i is descended; A_i is the cumulative additive effect of polygenes from the parental source i; D_{ij} is the cumulative dominance effect of polygenes from the mating of parental source i with parental source j, and $D_{ij}=D_{ji}$; C_i is the cumulative effect of cytoplasmic polygenes from parental source i; $\alpha(k)_i$, $\delta(k)_{ij}$ and $\gamma(k)_i$, which rest on the mating type of a genetic entry, are corresponding coefficients, respectively.

The genotypic value with respect to individual loci can be partitioned as:

$$Gs_{ijk} = \sum_s [\tau(k)_{ii} S_{sii} + \tau(k)_{jj} S_{sjj} + \tau(k_{ij}) S_{sij}], \tag{3}$$

where S_{sii}, S_{sjj} and S_{sij} are the genotypic effects at locus s from the combination of alleles from parent i and i, j and j, i and j, respectively; $\tau(k)_{ii}$, $\tau(k)_{jj}$ and $\tau(k)_{ij}$ are their coefficients, respectively. The coefficients of genetic components for parent, F_1, F_2, two backcrosses (BC_1 and BC_2) and two reciprocal backcrosses (RBC_1 and RBC_2) are shown in Table 1. In view of the generality (e.g. multiple allelism), genotypic values are adopted. If necessary, additive and dominance effects or other parameters at a given locus can easily be parameterized following the traditional line.

The total interaction effect between single genes and polygenes can be expressed as:

$$Gsp_{ijk} = \sum_s \left[\begin{array}{l} \sum\limits_{\substack{x \le y \\ w}} \tau(k)_{xy} \alpha(k)_w SA_{sxyw} + \sum\limits_{\substack{x \le y \\ w \le z}} \tau(k)_{xy} \delta(k)_{wz} SD_{sxywz} \\ + \sum\limits_{\substack{x \le y \\ w}} \tau(k)_{xy} \gamma(k)_w SC_{sxyw} \end{array} \right], \tag{4}$$

where $x, y, w, z \in \{i, j\}$; SA_{sxyw}, SD_{sxywz} and SC_{sxyw} are the interaction effects between the genotype at locus s made up of the allele from parent x and the allele from parent y, and additive effects of polygenes from parent w, dominance of polygenes from the mating of parent w and parent z, and cytoplasm of polygenes from parent w, respectively.

Correspondingly, interaction effects between genotype and environment can be similarly decomposed. The components of interaction effects between polygenes and environment are:

$$GpE_{hijk} = \sum_x \alpha(k)_x AE_{hx} + \sum_{x \le y} \delta(k)_{xy} DE_{hxy} + \sum_x \gamma(k)_x CE_{hx}, \tag{5}$$

where $x, y \in \{i, j\}$; AE_{hx}, DE_{hxy} and CE_{hx} are the interaction effects between the environment and additive, dominance, and cytoplasm effects respectively. Those of interaction effects between single genes and the environment are:

$$GsE_{hijk} = \sum_s [\tau(k)_{ii} SE_{hsii} + \tau(k)_{jj} SE_{hsjj} + \tau(k)_{ij} SE_{hsij}], \tag{6}$$

where SE_{hsii}, SE_{hsjj} and SE_{hsij} are the interaction effects between genotypes at locus s and the environment. Those of inter-

Table 1 Coefficients of genetic effects for commonly used mating types

k[a]	$\tau(k)_{ii}$	$\tau(k)_{jj}$	$\tau(k)_{ij}$	$\alpha(k)_i$	$\alpha(k)_j$	$\delta(k)_{ii}$	$\delta(k)_{jj}$	$\delta(k)_{ij}$	$\gamma(k)_i$	$\gamma(k)_j$
0	1	0	0	2	0	1	0	0	1	0
1	0	0	1	1	1	0	0	1	1	0
2	0.25	0.25	0.5	1	1	0.25	0.25	0.5	1	0
3	0.5	0	0.5	1.5	0.5	0.5	0	0.5	1	0
4	0	0.5	0.5	0.5	1.5	0	0.5	0.5	1	0
5	0.5	0	0.5	1.5	0.5	0.5	0	0.5	1	0
6	0	0.5	0.5	0.5	1.5	0	0.5	0.5	0	1

[a] Mating-type k designate: 0=Parent (P_i), 1=F_{1ij} ($P_i \times P_j$), 2=F_{2ij}, 3=BC_1 ($F_{1ij} \times P_i$), 4=BC_2 ($F_{1ij} \times P_j$), 5=RBC_1 ($P_i \times F_{1ij}$), 6=RBC_2 ($P_j \times F_{1ij}$), respectively

action effects between single genes, polygenes and environment are:

$$GspE_{hijk} = \sum_{s} \left[\begin{array}{l} \sum\limits_{\substack{x\le y \\ w}} \tau(k)_{xy}\alpha(k)_{w} SAE_{hsxyw} + \sum\limits_{\substack{x\le y \\ w\le z}} \tau(k)_{xy}\delta(k)_{wz} SDE_{hsxywz} \\ + \sum\limits_{\substack{x\le y \\ w}} \tau(k)_{xy}\gamma(k)_{w} SCE_{hsxyw} \end{array} \right], \quad (7)$$

where $x, y, w, z \in \{i, j\}$; SAE_{hsxyw}, SDE_{hsxywz} and SCE_{hsxyw} are the interaction effects of $S_{sxy}\times A_w\times E_h$, $S_{sxy}\times D_{wz}\times E_h$ and $S_{sxy}\times C_w\times E_h$, respectively.

Mixed linear model approaches can be used to analyze the above model. When the polygenic sources of parents are randomly sampled from a reference population, and the genotypes of parents with regard to single gene loci under consideration are known, locus effects can be treated as fixed effects and the polygenic effects and their interaction effects as random effects. In this situation, the genetic model can be rewritten in the following matrix form of a mixed linear model for all observations (written as the vector $\boldsymbol{y}$) in a modified diallel mating design:

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\mu + \sum_{s}\mathbf{X}_s\mathbf{b}_s + \mathbf{U}_E\mathbf{e}_E + \mathbf{U}_A\mathbf{e}_A + \mathbf{U}_D\mathbf{e}_D + \mathbf{U}_C\mathbf{e}_C \\ &\quad + \sum_{s}(\mathbf{U}_{SA_s}\mathbf{e}_{SA_s} + \mathbf{U}_{SD_s}\mathbf{e}_{SD_s} + \mathbf{U}_{SC_s}\mathbf{e}_{SC_s}) + \sum_{s}\mathbf{U}_{SE_s}\mathbf{e}_{SE_s} \\ &\quad + \mathbf{U}_{AE}\mathbf{e}_{AE} + \mathbf{U}_{DE}\mathbf{e}_{DE} + \mathbf{U}_{CE}\mathbf{e}_{CE} \\ &\quad + \sum_{s}(\mathbf{U}_{SAE_s}\mathbf{e}_{SAE_s} + \mathbf{U}_{SDE_s}\mathbf{e}_{SDE_s} + \mathbf{U}_{SCE_s}\mathbf{e}_{SCE_s}) + \mathbf{U}_B\mathbf{e}_B + \mathbf{I}\mathbf{e}_e \\ &= \mathbf{X}\mathbf{b} + \sum_{u}\mathbf{U}_u\mathbf{e}_u \end{aligned} \quad (8)$$

with the variance-covariance matrix

$$\begin{aligned} \mathrm{Var}(\mathbf{y}) &= \sigma_E^2\mathbf{U}_E\mathbf{U}_E^T + \sigma_A^2\mathbf{U}_A\mathbf{R}_A\mathbf{U}_A^T + \sigma_D^2\mathbf{U}_D\mathbf{R}_D\mathbf{U}_D^T + \sigma_C^2\mathbf{U}_C\mathbf{R}_C\mathbf{U}_C^T \\ &\quad + \sum_{s}(\sigma_{SA_s}^2\mathbf{U}_{SA_s}\mathbf{R}_{SA_s}\mathbf{U}_{SA_s}^T + \sigma_{SD_s}^2\mathbf{U}_{SD_s}\mathbf{R}_{SD_s}\mathbf{U}_{SD_s}^T \\ &\quad + \sigma_{SC_s}^2\mathbf{U}_{SC_s}\mathbf{R}_{SC_s}\mathbf{U}_{SC_s}^T) + \sum_{s}\sigma_{SE_s}^2\mathbf{U}_{SE_s}\mathbf{R}_{SE_s}\mathbf{U}_{SE_s}^T \\ &\quad + \sigma_{AE}^2\mathbf{U}_{AE}\mathbf{R}_{AE}\mathbf{U}_{AE}^T + \sigma_{DE}^2\mathbf{U}_{DE}\mathbf{R}_{DE}\mathbf{U}_{DE}^T + \sigma_{CE}^2\mathbf{U}_{CE}\mathbf{R}_{CE}\mathbf{U}_{CE}^T \\ &\quad + \sum_{s}(\sigma_{SAE_s}^2\mathbf{U}_{SAE_s}\mathbf{R}_{SAE_s}\mathbf{U}_{SAE_s}^T + \sigma_{SDE_s}^2\mathbf{U}_{SDE_s}\mathbf{R}_{SDE_s}\mathbf{U}_{SDE_s}^T \\ &\quad + \sigma_{SCE_s}^2\mathbf{U}_{SCE_s}\mathbf{R}_{SCE_s}\mathbf{U}_{SCE_s}^T) \\ &\quad + \sigma_B^2\mathbf{U}_B\mathbf{U}_B^T + \sigma_e^2\mathbf{I} \\ &= \sum_{u}\sigma_u^2\mathbf{U}_u\mathbf{R}_u\mathbf{U}_u^T = \mathbf{V} \end{aligned} \quad (9)$$

where **b** is the vector of fixed effects consisting of the mean and the effects of loci; X is the known incidence matrix relating to the fixed vector; $e_u \sim (0, \sigma_u^2 R_u)$ is the random vector for environment, additive, dominance, cytoplasm, locus *s*×additive, locus *s*×dominance, locus *s*×cytoplasm, locus *s*×environment, additive×environment, dominance×environment, cytoplasm×environment, locus *s*×additive×environment, locus *s*×dominance×environment, locus *s*×cytoplasm×environment, block and residual effects, respectively; R_u is the relative matrix for random vector e_u, R_u=I if the elements are independent, where I is a unit matrix; U_u is the known incidence matrix relating to random vector e_u; U_u^T is the transposition U_u; σ_u^2 is variance component of random effects e_u.

Variance components of the mixed linear model can be estimated by the methods of Henderson III (Henderson 1953; Searle 1968), minimum norm quadratic unbiased estimation (MINQUE) (Rao 1971), restriction maximum likelihood estimation (REML) (Paterson and Thomson 1971), and maximum likelihood estimation (ML) (Hartley and Rao 1967). Among these methods MINQUE possesses advantages of unbiasedness, no assumption of normal distribution, and less computation over REML and ML.

The estimates of MINQUE, REML, or ML asymptomatically conform to normal distribution, some classic statistical test methods (e.g. z-test, chi-square test) can be used for the significance test. The prediction of random effects can be obtained by the best linear unbiased prediction (BLUP) (Henderson 1963), linear unbiased prediction (LUP) (Zhu 1992; Zhu and Weir 1994), and adjusted linear unbiased prediction (AUP) (Zhu 1993; Zhu and Weir 1996) methods. Jackknife numerical resampling procedure (Miller 1974; Efron 1982) can also be used for estimating the sampling variance of estimated variance components and of predicted random effects, and Student's *t*-test for the significance test.

The fixed effects can be estimated through the Ordinary Least Squares (OLS or LS) method or the Generalized Least Squares (GLS) method. When the coefficient matrix of fixed effects X is degenerate, some jointly unestimatable constraints should be placed, or fixed effects should be re-parameterized for obtained a set of solutions with a biological sense, e.g. additive effects and dominance effects of single genes in convention are linear combinations of single gene genotypic values. The generalized linear test can be employed to test for the significance of parameters or estimable functions of parameters. The Jackknife technique can also be used in the significance test for fixed effects.

Results

Monte Carlo simulation results

Monte Carlo simulations were performed for two sets (Experiment I and Experiment II) of genetic entries in four environments with two randomized complete blocks from diallel designs of six parents, of which the polygenes originated from different sources; three were homozygous for one allele and the others were homozygous for another allele at an identifiable locus. The genetic mating design of Experiment I followed a complete diallel design which includes all parents, F_1s, and reciprocal F_1s while that of the Experiment II used a modified diallel design into which all F_2s and reciprocal F_2s are added. Since there were only two alleles at the single gene locus, the single gene effects were expressed as an additive effect (*a*) and a dominance effect (*d*) in accordance with conventional usage. In the case of two alleles, the additive effect determined from the comparison of two homozygotes and the dominance effect based on the comparison of the heterozygote with the average homozygote value, were parameterized. From Table 1, we can construct phenotypic values for parents *i* (*k*=0) by:

$$\begin{aligned} y_{hii0l} &= \mu + E_h + x_{ii}\,a + x_{ii}\,aE_h + 2A_i + D_{ii} + C_i + 2AE_{hi} + DE_{hii} + CE_{hi} \\ &\quad + x_{ii}(2aA_i + aD_{ii} + aC_i) + x_{ii}(2aAE_{hi} + aDE_{hii} + aCE_{hi}) \\ &\quad + B_{l(h)} + e_{hijkl}, \end{aligned}$$

where x_{ii} is an indicator variable which is 1 when *i* is a given allele and –1 when *i* is the other allele. Phenotypic values for F_{1ij} (*k*=1) from maternal parent *i*×paternal parent *j* are:

$$\begin{aligned} y_{hij1l} &= \mu + E_h + x_{ii}\,a + x_{ii}aE_h + A_i + A_j + D_{ij} + C_i + AE_{hi} + AE_{hj} + DE_{hij} \\ &\quad + CE_{hi} + x_{ii}(aA_i + aA_j + aD_{ij} + aC_i) \\ &\quad + x_{ii}(aAE_{hi} + aAE_{hj} + aDE_{hij} + aCE_{hi}) + B_{l(h)} + e_{hijkl} \end{aligned}$$

when two parents carry the same allele, or

$$\begin{aligned} y_{hij1l} &= \mu + E_h + d + dE_h + A_i + A_j + D_{ij} + C_i + AE_{hi} + AE_{hj} + DE_{hij} + CE_{hi} \\ &\quad + dA_i + dA_j + dD_{ij} + dC_i + dAE_{hi} + dAE_{hj} + dDE_{hij} + dCE_{hi} \\ &\quad + B_{l(h)} + e_{hijkl} \end{aligned}$$

Table 2 Bias and MSE of fixed effects estimated by OLS and GLS for Experiment I and Experiment II

Experiment	Parameter[a]	True value	OLS		GLS	
			Bias	MSE	Bias	MSE
I	μ	100	0.153	31.065	–0.161	30.264
	a	20	0.260	30.205	–0.021	21.657
	d	10	–0.025	10.157	0.171	10.483
II	μ	100	–0.110	32.529	–0.277	30.182
	a	20	0.214	53.825	–0.109	53.229
	d	10	0.189	59.350	–0.030	33.344

[a] μ, a, and d are mean, additive effect, and dominance effect, respectively

Table 3 Bias and MSE of variance components estimated by MINQUE (1) and Henderson III for Experiment I

Parameter[a]	True value	MINQUE (1)		Henderson III	
		Bias	MSE	Bias	MSE
σ^2_E	20	0.334	1,701.200	1.466	1,578.670
σ^2_A	20	–0.079	390.913	1.603	625.033
σ^2_D	20	0.375	98.583	–0.127	96.426
σ^2_C	20	0.826	334.383	0.931	262.996
σ^2_{aE}	20	–0.236	873.103	0.257	1,008.170
σ^2_{dE}	20	–0.098	405.255	0.722	444.107
σ^2_{AE}	20	0.268	91.621	0.086	116.689
σ^2_{DE}	20	0.263	27.800	–0.002	23.887
σ^2_{CE}	20	0.486	65.252	–0.718	53.352
σ^2_B	20	0.066	205.690	0.695	218.977
σ^2_e	10	0.024	1.109	–0.074	1.192

[a] Parameter represents variance component which σ^2_E=environment, σ^2_A=additive, σ^2_D=dominance, σ^2_C=cytoplasm, σ^2_{aE}=interaction between additive of single gene and environment, σ^2_{dE}=interaction between dominance of single gene and environment, σ^2_{AE}=interaction between additive of polygenes and environment, σ^2_{DE}=interaction between dominance of polygenes and environment, σ^2_{CE}=interaction between cytoplasm of polygenes and environment, σ^2_B=block, and σ^2_e=residual effect, respectively

when parents carry different alleles. And those of F_{2ij} (k=2) selfed from F_{1ij} are

$$\begin{aligned} y_{hij2l}=&\mu+E_h+x_{ii}a+x_{ii}aE_h+A_i+A_j+0.25D_{ii}+0.25D_{jj}\\ &+0.5D_{ij}+C_i+AE_{hi}+AE_{hj}+0.25DE_{hii}\\ &+0.25DE_{hjj}+0.5DE_{hij}+CE_{hi}\\ &+x_{ii}(aA_i+aA_j+0.25aD_{ii}+0.25aD_{jj}+0.5aD_{ij}+aC_i)\\ &+x_{ii}(aAE_{hi}+aAE_{hj}+0.25aDE_{hii}+0.25aDE_{hjj}\\ &+0.5aDE_{hij}+aCE_{hi})+B_{l(h)}+e_{hijkl} \end{aligned}$$

when two parents carry the same allele, or

$$\begin{aligned} y_{hij2l}=&\mu+E_h+0.5d+0.5dE_h+A_i+A_j+0.25D_{ii}+0.25D_{jj}\\ &+0.5D_{ij}+C_i+AE_{hi}+AE_{hj}+0.25DE_{hii}\\ &+0.25DE_{hjj}+0.5DE_{hij}+CE_{hi}\\ &+0.5(dA_i+dA_j+0.25dD_{ii}+0.25dD_{jj}+0.5dD_{ij}+dC_i)\\ &+0.5(dAE_{hi}+dAE_{hj}+0.25dDE_{hii}+0.25dDE_{hjj}\\ &+0.5dDE_{hij}+dCE_{hi})+B_{l(h)}+e_{hijkl} \end{aligned}$$

when parents carry different alleles.

Owing to some effects (e.g. the dominance effect of polygenes and its interaction with the single gene) are confounded, the simulations were based on a reduced genetic model without interaction effects between the single gene and polygenes in Experiment I. The parameter setting is shown in Tables 2, 3 and 4. Pseudo-random normal deviates with zero mean and unit variance were generated by the method of Kinderman and Monahan (1977). Henderson III and MINQUE (1) were used to estimate variance components while GLS and OLS were used to estimate fixed effects. For each case, 500 simulations were run to obtain sample means of the estimates, bias and Mean Squared Error (MSE).

Simulation results of bias and MSE are summarized in Table 2 for fixed effects and in Tables 3 and 4 for variance components. In the absolute value of bias is less than 5% of the parameter value, the estimate was regarded as unbiased. The largest absolute value of bias was 1.89% in Table 2. As indicated, fixed effects could be unbiasedly estimated by both OLS and GLS. The MSE of fixed effects, which reflected the variation of estimates among different simulations, were small while GLS had a little smaller MSE than OLS. The results in Table 2 suggested that both estimates of OLS and GLS are unbiased and efficient for additive and dominance effects of single genes in both experimental designs. Although the results of simulations showed GLS might be a little more efficient, it requires the inversion of a variance-covariance matrix and involves heavy computations, especially in cases with a large set of data. Hence, it is worth employing the OLS method

Table 4 Bias and MSE of variance components estimated by MINQUE (1) and Henderson III for Experiment II

Parameter[a]	True value	MINQUE (1)		Henderson III	
		Bias	MSE	Bias	MSE
σ^2_E	20	–0.240	1,557.810	–0.139	1,792.960
σ^2_A	20	–0.1945	381.380	–1.517	906.664
σ^2_D	20	0.313	372.160	17.431	1,444.730
σ^2_C	20	0.393	266.323	0.010	383.117
σ^2_{aA}	20	–1.287	565.514	3.253	672.820
σ^2_{aD}	20	1.247	1,512.563	–20.000	400.000
σ^2_{aC}	20	0.708	442.578	1.828	828.233
σ^2_{dA}	20	0.391	1,510.360	–20.000	400.000
σ^2_{dD}	20	0.726	720.846	–2.431	738.441
σ^2_{dC}	20	–1.393	320.631	–1.327	384.140
σ^2_{aE}	20	–0.680	2,324.125	2.010	2,677.280
σ^2_{dE}	20	–0.654	1,674.890	10.014	2,852.130
σ^2_{AE}	20	–0.455	99.665	–1.035	172.502
σ^2_{DE}	20	0.009	91.115	15.002	505.266
σ^2_{CE}	20	–0.268	65.682	0.184	80.590
σ^2_{aAE}	20	1.291	154.405	5.632	277.192
σ^2_{aDE}	20	–1.057	304.670	–20.000	400.000
σ^2_{aCE}	20	0.503	135.785	–0.916	257.818
σ^2_{dAE}	20	0.732	342.910	–20.000	400.000
σ^2_{dDE}	20	–0.297	527.826	0.246	764.977
σ^2_{dCE}	20	0.924	165.566	–0.900	255.138
σ^2_B	20	0.864	202.227	–0.387	206.043
σ^2_e	10	–0.037	0.528	0.031	0.575

[a] Parameter represents variance component which σ^2_E, σ^2_A, σ^2_D, σ^2_C, σ^2_{aE}, σ^2_{dE}, σ^2_{AE}, σ^2_{DE}, σ^2_{CE}, σ^2_B and σ^2_e are the same as Table 3; σ^2_{aA}=additive-additive epistasis between single gene and polygenes, σ^2_{aD}=additive-dominance epistasis, σ^2_{aC}=additive-cytoplasm epistasis, σ^2_{dA}=dominance-additive epistasis between single gene and polygenes, σ^2_{dD}=dominance-dominance epistasis, σ^2_{dC}=dominance-cytoplasm epistasis, σ^2_{aAE}=interaction between additive-additive epistasis and environment, σ^2_{aDE}=interaction between additive-dominance epistasis and environment, σ^2_{aCE}=interaction between additive-cytoplasm epistasis and environment, σ^2_{dAE}=interaction between dominance-additive epistasis and environment, σ^2_{dDE}=interaction between dominance-dominance epistasis and environment, σ^2_{dCE}=interaction between dominance-cytoplasm epistasis and environment, respectively

and losing a little precision for the sake of saving the calculation.

All the bias of estimated variances approached to zero by using MINQUE (1), while two absolute values of bias were larger than 5% of the true values by using Henderson III in Experiment I (Table 3), MSE of Henderson III were similar to those of MINQUE (1). Most of MSE were small enough. But a few MSE, e.g. of σ^2_Es and σ^2_{aE}s, were larger, which might result from small sample size (only four for the environment) and/or complication of the genetic model, and might largely reduce by augmentation of sample size and simulation replications. It was suggested that MINQUE (1) is unbiased and efficient for estimating the variance component and Henderson III is almost as good as MINQUE (1) in Experiment I.

For Experiment II (Table 4), Henderson III estimation resulted in several large biases (for σ^2_D, σ^2_{dE} and σ^2_{DE}) and inestimatable parameters (of σ^2_{aD}, σ^2_{dA}, σ^2_{aDE} and σ^2_{dAE}), which were due to completely linear dependency of the coefficient matrix. Therefore, Henderson III is not robust and/or is inapplicable with the model being complicated. But MINQUE (1) could almost hold a similar estimation precision and efficiency to those in Experiment I. From the results of simulations, it was suggested that MINQUE (1) was unbiased and efficient in both experiments, with Henderson III only in Experiment I, for estimating variance components.

Worked example

The data from Gilbert(1985a) were used to illustrate the use of a reduced genetic model and corresponding methods for estimations of fixed effects and variance components. There is an allozyme locus, i.e. *esterase 6*, encoding an enzyme in which different migration rates for alleles can be detected by the electrophoretic technique in *D. melanogaster.* Putative functions of the esterase 6 enzyme include male reproduction and larval nutrition. Six strains derived from natural population, three of which were homozygous for *Est* 6^{S2} and the others homozygous for *Est* 6^{F2}, were used to produce all possible progeny following a 6×6 complete diallel design. Mating speed, copulation duration, developmental time, and enzyme activity were measured. An ANOVA-based analysis was

Table 5 ANOVA table for mating speed, copulation duration, developmental time, and esterase 6 activity in *D. melanogaster*

Source	*df*	Mating speed (√ min/male)		Copulation duration (min/male)		Developmental time (h-200/fly)		Esterase 6 activity 10^{-8}M β-napthol/male	
		SS	MS	SS	MS	SS	MS	SS	MS
Additive locus	1	2.993	2.993	22.222	22.222	0.273	0.273	150.222	150.222
Additive strain	4	0.477	0.119	210.232	52.558	1,725.495	431.374	29,840.92	7,460.231
Dominant locus	1	0.134	0.134	0.238	0.238	129.753	129.753	186.166	186.166
Dominant strain	14	7.42	0.53	50.112	3.579	1,243.865	88.847	12,059.35	861.382
Maternal effect	5	0.322	0.064	4.834	0.967	2,193.671	438.734	30,476.72	6,095.345
Remainder	10	1.318	0.132	18.666	1.867	865.272	86.527	11,250.28	1,125.028

Table 6 Estimates of parameters for mating speed, copula duration, developmental time, and esterase 6 activity in *D. melanogaster*

Parameter[a]	Mating speed[b]	Copulation duration	Developmental time	Esterase 6 activity
μ	2.77***	19.11***	79.43***	131.06***
a	0.47**	–1.20	3.07	–9.43
d	–0.19	–0.11	3.81	–4.56
σ_A	0.000	2.040$^+$	0.000	11.619
σ_D	0.480*	1.000	1.162	0.000
σ_M	0.000	0.000	10.817*	40.743*
σ_e	0.361	1.367	9.301	33.615

[a] μ, a, and d are mean, additive effect, and dominance effect, respectively; σ_A, σ_D, σ_M, and σ_e are standard deviations of additive effect, dominance effect, maternal effect, and residual effect, respectively

[b] Level of significance: $^+$ $P<0.10$, * $P<0.05$, ** $P<0.01$, *** P<0.001 respectively

conducted by Gilbert (1985a). The data were re-analyzed by the present methods. Maternal effect was confounded with cytoplasmic effect and they cannot be distinguished in Gilbert's diallel design. Thus, the designation of maternal effect was here retained, but it was, more pertinently, a maternal nuclear effect plus a cytoplasmic effect. Calculation of these was equivalent to that of cytoplasmic effects. The estimates of ANOVA were used to estimate single gene effects in Gilbert's paper, but they are not linear and unbiased. For example, $\hat{a}^2=(MS1-MS2)/2m^2$, and $|\hat{a}|=\sqrt{MS1-MS2}/2m^2$. This can be guaranteed for $E(\hat{a}^2)=a^2$, but not for $E(|\hat{a}|)=|a|$, i.e. an unbiased quadric form cannot guarantee that the first oder mean is also unbiased. Fixed effects were therefore here estimated by GLS. As Henderson III is unbiased and efficient in the case of the reduced genetic model, this method was employed to estimate variance components. The results of ANOVA were the same as those of Gilbert (see Table 5). Estimates of fixed effects and variance components were presented in Table 6.

Results showed that the additive effect of the *esterase 6* locus was significant for mating speed but non-significant for the other traits. No significant dominance effect of the locus was detected on all four traits. Variation from polygenes among strains could be measured by estimates of variance components. Maternal effects were significant for developmental time and esterase 6 activity. Variances of additive polygene effects on copulation duration and esterase 6 activity and of dominant polygene effects on mating time, copulation duration, and developmental time were larger than zero, but whether significant or not could not be determined since the F test could not be used. An approximate test based on the asymptotic normal distribution of estimates showed that the dominant variance of mating time was significant at the 0.05 level while additive variance was significant at the 0.10 level.

Discussion

Both major genes and polygenes are responsible for the genetic differences in many quantitative traits. In practice, the exploitation of both major genes and polygenes was included in some breeding schemes to improve important traits, e.g. dwarfish plant, disease resistance, grain quality, etc. The effects of major genes and polygenes often follow different genetic models. This is the basis of not only expounding their gene expressions but also developing an efficient breeding strategy to correctly dissect genetic effects in terms of different variation sources. Exact separation of them rests on the application of proper analysis methods. It is therefore crucial to develop methods suited to these tasks. Based on this, those polygenes with sufficiently small effects in relation to total variation are numerous enough to warrant the continuous and normal distribution of the genotypic values of individuals in the population under study; some methods were developed for simultaneously estimating the variation components of both types of genes. If major

genes are identifiable or distinguishable, the ANOVA methods of Gilbert (1985a, b) and Elkind and Cahaner (1986), etc., can be used. Whereas, ANOVA requires data with a balanced structure, it cannot handle genetic designs with irregular missing combinations and complicated genetic models. The mixed linear model approach applied in the present research is a powerful statistical tool for tackling those difficult problems since it can almost deal with data with an arbitrary structure.

Usually, generalized conclusions need to be drawn on genetic effects in the reference population studied. Diallel analysis can provide much information about the genetic features and properties of polygenes in parents from a sampled population. Diallel designs were extensively used in genetics and in breeding for this purpose. If the sources of polygenes are randomly derived from a natural population under Hardy-Weinberg equilibrium and linkage equilibrium, diallel analysis permits a valid estimation of genetic parameters in that population. When, however, variation of the metric trait is contributed both by major genes and by polygenes, traditional diallel analysis cannot provide the right and unequivocal information about their genetic effects. Gilbert (1985a, b) put forward an enlightened idea to estimate single gene effects as well as genetic effects of polygenes on quantitative traits from a diallel design, although some limitations existed in his method. The present genetic model and methodology is developed mainly for diallel designs or modified diallel designs which contain progenies derived from them. Meanwhile, some improvement have been made on Gilbert's method, e.g. unbiasedness for estimates of single gene effects, capability of handling unbalanced data, interaction effects between single genes and polygenes, and more than one locus, etc. Moreover, owing to the generality of the mixed linear model approach and flexibility of the genetic model, this new methodology can be directly applied to, or easily extended to, genetic entries other than those of the diallel mating design. By making some changes it can also be applied to the genetic analysis of other genetic materials, e.g. RIL populations, DH populations, and progeny families derived from a cross between two parents.

Some widely used methods, e.g. the isogenic-lines method and the diallel method, obtain estimates of the genetic effects of major genes and polygenes without considering the interaction between such genes. Thus, reliability of those conclusions lies on the nonexistence of the interaction. Since evidence of such an interaction has been reported (Elkind and Cahaner 1986; Jiang et al. 1994; Jiang and Liu 1995; Jiang et al. 1995), an appropriate method is needed to check wheter the interaction exists or not. The interaction effects can be analyzed if F_2, BC, and/or other subsequent generations, are added into genetic mating designs. When an experiment includes only parents and F_1, the method of estimating interaction effects is to add some lines into a mating scheme derived from the same ancestry or a shared common source of polygenes by descent, e.g. isogenic lines. On the other hand, no interactions between single genes and between polygenes are taken into account in the present research. When these really exist, the genetic model can readily be extended following Cockerham's (1980) approaches for polygenic effects and directly using the aggregate genotypic effects attributable to several given loci, or fitting digenic and/or higher-order epistasis terms to the model for single gene effects.

The interest of the present research focused on the estimation of single gene effects and the variances of polygenes when single genes can be scored or inferred from a pedigree. If major-gene genotypes cannot be known, the ML method based on mixture distribution can be used to identify the genetic model and to estimate parameters (Jiang et al. 1994; Jiang and Mo 1995; Jiang and Liu 1995; Jiang et al. 1995; Wang and Gai 1997; Gai and Wang 1998). However, those methods are only applicable to the case in which the observation is a mixture of finite component distributions such that $\mathbf{y}_i \sim MVN(\mathbf{1}\mu_i, \sigma_i^2\mathbf{I})$, but are not applicable to the more general situation in which $\mathbf{y}_i \sim MVN(\mathbf{X}_i\mathbf{b}_i, \mathbf{V}_i)$. Therefore, the estimation of interaction effects between major genes and polygenes and between polygenes and environment is an intractable problem. Some researchers (Jiang et al. 1995) declared that the inequality of variances of polygenes with different major-gene genotypes in the same families indicated the existence of interaction between major genes and polygenes. But attention must be paid to the fact that equality of those variances may not be due to the absence of interaction, just like the case illustrated in Fig. 2B in Elkind and Cahaner's paper (1986). Appropriate analysis methods can thus be developed through using mixed linear-model approaches.

Acknowledgements This work was supported by the Transcentury Training Program Foundation for the Talents of the National Education Commission of PRC, National Natural Scientific Foundation of PRC, and Foundation of Education Committee of Zhejiang Province. The experiments comply with the current laws of the country in which the experiments were performed.

References

Cockerham CC (1980) Random and fixed effects in plant genetics. Theor Appl Genet 56:119–131

Comstock RE, Robinson HF (1952) Estimation of average dominance of genes. In: Gowan JW (ed) Heterosis. Iowa State University Press, Ames, Iowa, pp 494–516

Eberhart SA, Gardner CO (1966) A general model for genetic effects. Biometrics 22:864–881

Efron B (1982) The Jackknife, the Boostrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia

Elkind Y, Cahaner A (1986) A mixed model for the effects of single gene, polygenes and their interaction on quantitative traits. 1. The model and experimental design. Theor Appl Genet 72:377–383

Falconer DS (1981) Introduction to quantitative genetics (2nd edn.). Longman House, London

Fernando RL, Stricker C, Elston RC (1994) The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. Theor Appl Genet 88:573–580

Gai JY, Wang JK (1998) Identification of major gene and polygene mixed inheritance model from backcrosses or $F_{2:3}$ families. Acta Agron Sinica 24:402–409

Gardner CO, Eberhart SA (1966) Analysis and interpretation of the variety cross diallel and related populations. Biometrics 22:439–452

Gilbert DG (1985a) Estimating single gene effects on quantitative traits. I. A diallel method applied to *Est 6* in *D. melanogaster.* Theor Appl Genet 69:625–629

Gilbert DG (1985b) Estimating single gene effects on quantitative traits. II. Statistical properties of five experimental methods. Theor Appl Genet 69:631–636

Gilbert NEG (1958) Diallel cross in plant breeding. Heredity 12: 477-492

Griffing BI (1956) Concept of general and specific combining in relation to diallel crossing systems. Aust J Biol Sci 9:463–493

Hartley HO, Rao JNR (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrics 54:93–108

Hayman BI (1954a) The analysis of variance of diallel tables. Biometrics 10:235–244

Hayman BI (1954b) The theory and analysis of diallel cross. Genetics 39:789–809

Henderson CR (1953) Estimation of variance and covariance components. Biometrics 9:226–253

Henderson CR (1963) Selection index and expected genetic advance. In: Hanson WD, Robinson HE (eds) Statistical genetics and plant breeding. Natl Acad Sci USA 982: pp 141–163

Jiang CJ, Liu XF (1995) Genetic analysis of quantitative trait differences with effects of major genes among inbred lines. Acta Genet Sinica 22:59–64

Jiang CJ, Mo HD (1995) Genetic analysis for qualitative-quantiative traits. IV. Application of the maximum likelihood method. Acta Agron Sinica 19:641–648

Jiang CJ, Pan XB, Gu MH (1994) The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. Genetics 136:383–394

Jiang CJ, Xu CW, Hui DF, Shao YJ (1995) Analysis of the effects of major genes on quantitative traits between families. Acta Agron Simica 21:632–636

Kinderman AJ, Monahan JF (1977) Computer generation of random variables using the ratio of normal derivates. Assoc Comput Mach Trans Math Software 3:257–260

Liosel P, Goffinet B, Monod H, Oca GMD (1994) Detecting a major gene in an F_2 population. Biometrics 50:512–516

Mather K, Jinks JL (1982) Biometrical genetics (3rd edn). Chapman and Hall, London

Miller RG (1974) The Jackknife: a review. Biometrika 61:1–15

Mo HD (1993a) Genetic analysis for qualitative-quantitative traits. I. The genetic constitution of generation populations and identification of major gene genotypes. Acta Agron Sinica 19: 1–6

Mo HD (1993b) Genetic analysis for qualitative-quantitative traits. II. Generation means and genetic variances. Acta Agron Sinica 19:193–200

Mo HD, Xu CW (1994) Genetic analysis for quanlitative-quantitative traits. III. Endosperm characters under triploid genetic control. Acta Agron Sinica 20:513–519

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biomatrika 58:545–554

Rao CR (1971) Estimation of variance and covariance components: MINQUE theory. J Multivar Anal 1:257–275

Searle SR (1968) Another look at Henderson's methods of estimating variance components. Biometrics 24:749–778

Wang JK, Gai JY (1997) Identification of major gene and polygene mixed inheritance model and estimation of genetic parameters of a quantitative trait from F_2 progeny. Acta Genet Sinica 24:432–440

Yates F (1947) Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity 1:287–301

Zhu J (1992) Mixed model approaches for estimating genetic variances and covariances. J Biomath 7:1–11

Zhu J (1993) Mixed model approaches for estimating genetic covariances between two traits with unequal design matrices. J Biomath 8:24–30

Zhu J, Weir BS (1994) Analysis of cytoplasmic and maternal effects. II. genetic models for triploid endosperms. Theor Appl Genet 89:160–166

Zhu J, Weir BS (1996) Diallel analysis for sex-linked and maternal effects. Theor Appl Genet 92:1–9